

SUMEX

STANFORD UNIVERSITY
MEDICAL EXPERIMENTAL COMPUTER RESOURCE

RR - 00785

COMPETING RENEWAL APPLICATION

Submitted to

BIOTECHNOLOGY RESOURCES PROGRAM
NATIONAL INSTITUTES OF HEALTH

and Ann. Rpt. 07.

June 1, 1980

STANFORD UNIVERSITY SCHOOL OF MEDICINE
Edward A. Feigenbaum, Principal Investigator

Prelude: An Overview and Personal Statement

by Edward A. Feigenbaum, Principal Investigator

This prelude is unabashedly a statement of advocacy. As we prepare this proposal, gathering up the threads of our past achievement and weaving them into a coherent picture of our future, there is in the SUMEX Project a sense of pride and accomplishment, and a feeling of exhilaration and momentum regarding the future.

SUMEX was established with three main goals:

1. to provide computing resources and human assistance to those scientists working on applications of artificial intelligence research in medicine and biology;
2. to test the idea that it was feasible to provide resources and assistance to the nation from a single site, with time-shared operating systems, national computer communication networks, and a staff oriented toward the special problems of remote users;
3. to grow, from seed to plant, the community of scientists interested in working on applications of AI to the biomedical sciences; facilitating the growth, health, and vigor of the community by the use of electronic communications linking its members. One question we were asking was, "Is there a new style of science that will emerge in a communications-enhanced setting of national, rather than institutional, scope?"

These goals were and are unique to SUMEX, and their pursuit has given rise to a "spirit of SUMEX"--a spirit that unfortunately does not come across well in the dry recitations of a proposal document; hence, this personal prelude.

SUMEX's success as a national research resource

The SUMEX Project has demonstrated that it is possible to operate a computing research resource with a national charter--that the services providable over networks were those that were facilitative of the growth of AI-in-Medicine. Previous NIH computer RR's were mostly institutional in scope, occasionally regional (like the UCLA resource).

Some of the most notable projects in the history of Artificial Intelligence were done with terminal-and-network, without a computer on site. In human terms, this means, of course, without the headaches and energy drains of proposing a machine, installing it, maintaining it and its software, hiring its system programmers and operators, dealing with communication vendors, etc. The famous INTERNIST program was developed from Pittsburgh in this way. And the ACT computer model was begun at

Michigan, continued at Yale, and later at Carnegie-Mellon, all without moving the program or losing a day's work because of machine transition problems.

The projects SUMEX supports have generally required substantial computing resources with excellent interaction. This is hard to obtain in all but a few universities. SUMEX is, in a sense, a "great equalizer". A scientist gains access by virtue of the quality of his/her research ideas, not by the accident of where s/he happens to be situated--in other words, the ethic of the scientific journal.

SUMEX has demonstrated that a computer resource is a useful "linking mechanism" for bringing together and holding together teams of experts from different disciplines who share a common problem focus. For example, computer scientists have been collaborating fruitfully with physical chemists, molecular biochemists, geneticists, crystallographers, internists, ophthalmologists, infectious disease specialists, intensive care specialists, oncologists, psychologists, biomedical engineers, and other expert practitioners. And in some of these cases, the interdisciplinary collaboration, usually so difficult to achieve in the best of circumstances, was achieved in spite of geographical distance between the participants, using the computer networks.

SUMEX has achieved successes as a community builder. AI concepts and software are among the most complex products of computer science. Historically it has not been easy for scientists in other fields to gain access to and mastery of them. Yet the collaborative outreach of SUMEX has been able to bridge the gap in a number of cases. For example, Dr John Osborn (Pacific Medical Center, San Francisco) and I found common scientific interests in the application of AI to intensive care, and initiated a SUMEX-based collaboration. That project resulted in a system of potential significance to intensive care medicine; in two Stanford computer science Ph.D. dissertations, hence two new doctoral-level recruits to the ranks of computers-in-medicine specialists; in one computer science/physiology Special Ph.D. Program for one of Dr. Osborn's biomedical engineers; and an award to Dr. Osborn's team in 1979 from the Association for the Advancement of Medical Instrumentation.

I wish to contrast this success story with the traditional difficulties I have encountered outside the health research field in trying to bridge the gap to engineering-oriented industrial firms. The human resource and motivation was present. The SUMEX base of easily available shared software technology was not. The resulting problems have generally raised too high a threshold to overcome.

The SUMEX mission has been able to capture the contributions of some of the finest computers-in-medicine specialists and computer scientists in the country. For example, Professor Joshua Lederberg (SUMEX's first PI, now President of The Rockefeller University) is Chairman of SUMEX's Executive Committee; and Professor Donald Lindberg, M.D., Director of the University of Missouri's Health Care Technology Center, is Chairman of the AIM Advisory Group. Professor Herbert Simon of Carnegie-Mellon University, Professor Marvin Minsky of MIT, and many other distinguished scientists

serve on that peer review committee. These people are active participants in SUMEX. Lederberg and Lindberg are continuing collaborators in the research itself. And Simon, for example, was the person who prompted our collaboration with psychologists at the University of Colorado.

SUMEX now has the reputation of a model national resource, pulling together the best available interactive computing technology, software, and computer communications in the service of a national scientific community. Planning groups for national facilities in cognitive science, computer science, and biomathematical modeling have discussed and studied the SUMEX model.

SUMEX and Artificial Intelligence Research

The SUMEX Project is a relative latecomer to AI research. Yet its scope has given strong impetus to this historic development in computer application. AI research is that part of computer science that investigates symbolic reasoning processes, and the representation of symbolic knowledge for use in inference. It views heuristic knowledge to be of equal importance with "factual" knowledge, indeed to be the essence of what we call "expertise". In its "Expert Systems" work, it seeks to capture the expertise of a field, and translate it into programs that will offer intelligent assistance to a practitioner in that field.

For computer applications in medicine and biology, this research path is crucial, indeed ineluctable. Medicine and biology are not presently mathematically-based sciences; not like physics and engineering capable of exploiting the mathematical characteristics of computation. They are essentially inferential, not calculational, sciences. If the computer revolution is to affect biomedical scientists, computers will be used as inferential aids.

Perhaps the larger impact on medicine and biology will be the exposure and refinement of the hitherto largely private heuristic knowledge of the experts of the various fields studied. The ethic of science that calls for the public exposure and criticism of knowledge has traditionally been flawed for want of a methodology to evoke and give form to the heuristic knowledge of scientists. The AI methodology is beginning to fill that need. Heuristic knowledge can be elicited, studied, critiqued by peers, and taught to students.

The tide of AI research and application is rising. AI is one of the fronts along which university computer science groups are expanding. The NSF's program in Intelligent Systems is vigorous and growing. The pressure from student career-line choices is great: to cite an admittedly special case, approximately one-third of the students applying to Stanford's computer science Ph.D. program cite AI as a possible field of specialization. In industry, new groups have been forming regularly: Texas Instruments two years ago formed a substantial AI group; so did the oil-industry-service firm, Schlumberger, Inc.; IBM has reinitiated its AI work; and the new genetic engineering firms are becoming interested.

The tide is rising largely because of the development in the 1970's of methods and tools for the application of AI concepts to difficult professional-level problem solving; and the demonstration in various areas of medicine and other life sciences that these methods and tools really work. Here SUMEX has played a key role, so much so that it is regarded as "the home of applied AI."

SUMEX has been the nursery, as well as the home, of such well-known AI systems as DENDRAL (chemical structure elucidation), MYCIN (infectious disease diagnosis and therapy), INTERNIST (differential diagnosis), and ACT (human memory organization). These, and other programs developed at SUMEX, have played a seminal role in structuring modern AI paradigms and methodology. First among these has been a shift of AI's focus from inference procedures to knowledge representation and use. There is now a recognition that the power of problem solvers derives primarily from the knowledge that they contain--of the elements of the problem domain, of the strategies for solving problems in that domain, and of the forms in which the knowledge is to be acquired. In 1977, Goldstein and Papert of MIT, writing in the journal *Cognitive Science*, described the change of focus as a "paradigm shift" in AI. This shift was induced largely (though of course not exclusively) by the work at SUMEX, beginning with the DENDRAL development in 1965.

Toward the mid-'80s: the Future of SUMEX

Success breeds its problems. The revolution in computer technology and costs adds complexity to their solution.

At the beginning, the SUMEX community was small, and idea-limited. The SUMEX computer facility was an ideal vehicle for the research. Now the community is large, and the momentum of the science is such that its progress is now limited by computing power. The size and scientific maturity of the SUMEX community has fully consumed the resource in every critical dimension: CPU power, main memory size, and file space.

The limitation that AI researchers agree most critically limits their scientific imagination, and adds inordinately to program development time, is the 256K word main memory space, brought about by the 18 bit address of the PDP-10's and 20's. Economically, main memory size need not be much of a limitation any more, but it is essential to move to a machine with more addressing bits.

But which machine? In the turmoil of the computer developments of today, this is not easy to answer. Computers will come in many different sizes and prices and each will fit a particular class of needs. Our planning axiom for the period 1981-86 has been: the need to accommodate a HETEROGENEITY of computers and peripheral devices. We must maintain a flexible posture with respect to the introduction of new capabilities and changing costs during this continuing revolution. Yet we must choose.

Our plan, sketched below, is conservative in maintaining and extending SUMEX's current service level; yet is forward-looking enough to

position SUMEX properly for mid-course corrections and for the computing world of the late 1980's. Here it is, briefly sketched.

The existing DEC KI-10 duplex, with its superb software, will be "filled out"--stretched to the point of diminishing returns from hardware addition; then frozen. It is an amiable workhorse. We can not (indeed dare not) do without it during this period of turbulence. But it has seen better days, and will be ineffective by the end of the grant period.

A DEC VAX 11/780 will be acquired in the first year. Based on more modern technology and a more competitive price, it has the extra address bits that are required. On VAX we get the same kind of low-cost ride on the software work of others that we got when we adopted TENEX and INTERLISP for the KI-10's. The UNIX operating system is available, and is being further developed under ARPA support. ARPA is also supporting the reprogramming of INTERLISP for VAX. For integrated circuit design research, ARPA has already placed two VAX computers at our Computer Science Department, so we are building experience rapidly in VAX use. And, de facto, the VAX has become the "computer science machine" of the early '80s, so that nationally its software development is moving rapidly. A family of VAX's, both more and less powerful, at (hopefully) appropriate prices, is in the wings.

The "technology transfer" machine to which we will move the heavy national use of SUMEX's mature AI applications (such as DENDRAL, SECS, MOLGEN, VM) will be another DEC VAX, acquired in the middle of the period. This machine's role is intended to be entirely analogous to the role currently played by the DEC 2020 at SUMEX vis a vis the KI-10 duplex. It will be the VAX-era prototype of the "spinoff" machine, loosely tethered to SUMEX by networks. In the last DENDRAL Project renewal, the NIH Study Section denied such a machine to DENDRAL, suggesting that the required resource would better be provided by SUMEX. We seek, and plan, to assume this obligation.

And what about the single-user professional scientific workstation--the powerful, small, cheap officemate that will serve most of the researcher's computing needs? Much of the present turbulence in the computing world swirls around this question. Yes, we believe it is coming, and will probably be an economically viable concept in the late '80s. No, we do not believe it will be powerful enough or cheap enough for most routine research needs in the planning period. Yet we must begin to explore the space of possibilities opened up by these machines, eschewing articles of faith for real experience. We must learn to build systems of these machines and to build and manage graceful software for these systems. If decentralization is in our future, we must learn its technical characteristics. Consequently, we have planned the acquisition of a number of such single-user workstations over the course of the coming period, some to be placed at Stanford, some in the national community, at the decision of the Executive Committee.

These machines will be tethered to the SUMEX central facility and staff by local digital network at Stanford and by national network to the non-Stanford community. With DEC 10's, 20, VAX's, and workstations

coexisting to serve community needs, it is economical and convenient to continue the centralization of file storage, and the networks make it possible for most applications at Stanford and many applications nationally. Computer scientists are in general agreement that economies of scale will continue to dominate in secondary storage for some time. We have planned, therefore, to alleviate the present file space shortage not by adding discs to machines in an ad hoc fashion but by adding a common file server to the resource. To facilitate the transfer of software and access to valuable common facilities, the SUMEX complement of equipment will be linked by local digital networks to other major centers of computing at Stanford, most important of which is the Computer Science Department.

The success of SUMEX is the success of its dedicated and extraordinarily competent staff, headed by Tom Rindfleisch. This human resource of SUMEX should not, and will not, be decentralized. In the world of computer systems talent and user-assistance expertise, there are indeed continuing large "economies of scale".

The smoothly operating management structure of SUMEX is one of its joys and victories. We do not plan to fix something that is not broken. We plan that the Executive Committee and the AIM Advisory Committee will continue to function as they now do.

So this is it in a nutshell:

Run the present configuration with more main memory; acquire two VAX large-memory systems (years 1 and 3) for new research and for maturing project communities; cautiously add some single-user professional workstations; acquire a common file server; link everything in a transparent digital networking scheme; continue the central staff and management structure, essentially unchanged in size and function.

As we add up the budget (flinchingly, I hasten to say), we note that the cost will not be cheap, despite the much-touted fall in the cost of computing. But we believe we have been conservative; that the scientific community we serve needs these resources; and that by its science and its applications orientation, it has earned them.

I look at the widely acclaimed NSF report calling for the refurbishing of computer equipment for experimental computer science (the so-called "Feldman Report") and note that it calls for "refurbishing" expenditures for just a single department greater than that budgeted in this proposal, with a "refresh" cycle of five years to accommodate advancing technology. The scientific work of the SUMEX-AIM community is the quintessence of experimental computer science. It is advancing, and gaining acceptance, beyond expectations. SUMEX serves the nation, not one university or department. I believe that its budget accords well with the national interest and with the scientific interest.

Conclusion: the "Spirit of SUMEX"

I would like to conclude not with my own words but with the words of Professor Douglas Brutlag, a Stanford Biochemist who collaborates with my group on the MOLGEN project and who sent me, unsolicited, the letter quoted below in its entirety. Nothing I could say could more accurately portray the "spirit of SUMEX" mentioned earlier.

"My original role in the Molgen project was that of a biochemist advisor to those developing a knowledge base of molecular biological information and techniques. I rapidly found that SUMEX could be very useful to my own work in ways that I had never expected. First, MOLGEN was a success very early and I now routinely use the artificial intelligence methods incorporated within the frame oriented knowledge base in my everyday work in the laboratory. I use the knowledge base not only to store our results from experiments and to analyze them, but I can readily interact with the knowledge base to examine the data from several different viewpoints and display it in different ways.

In addition to the interactive nature of knowledge base work, I have found computer networks and file transfer protocols to be exceptionally useful. The nation wide commercial networks have permitted many of my colleagues across the country to try out the software we have developed at Stanford in pilot projects. This together with message sending capabilities has resulted in instantaneous feed back about the work we have done and allowed us to develop our program and to incorporate ideas from a much larger base of expertise. Several collaborative arrangements have been set up and some have even become involved in our programming efforts. Moreover, our software has had such general utility that subsequently many of the other workers have obtained accounts on their local computers and we have sent them the software by file transfer protocols. Electronic information transfers have saved both time and energy in preparing hard copy versions as well as facilitated the update programs at many distant locations.

I think that one of the major reasons that SUMEX works so well is that it is designed with the naive user in mind. Because it is so interactive and user oriented, the activation energy to learn how to use the system is very low. Of all of the interactive systems with which I have worked (five in all), SUMEX was not only the easiest, but was indeed a real pleasure. I felt more like the system was working for me from the very beginning, rather than me fighting the system. Hence, my productivity on SUMEX has increased immeasurably. In addition, I have no hesitation encouraging others at remote sites to use SUMEX in the collaborative efforts mentioned above."

Table of Contents

Section	Page
Prelude: An Overview and Personal Statement	i
List of Figures	xiii
1. Biographical Sketches	2
2. Budget	3
2.1 First Year Budget Detail (8/1/81 - 7/31/82)	3
2.1.1 Total First Year Budget	3
2.1.2 First Year Personnel Detail	4
2.2 5-year Budget Summary (8/81 - 7/86)	5
2.3 Budget Explanation and Justification	6
3. Introduction and Aims	13
3.1 Overview of Objectives and Rationale	14
3.1.1 Definitions of Artificial Intelligence	14
3.1.2 Resource Sharing	16
3.2 SUMEX-AIM Background	16
3.3 Specific Aims	18
3.3.1 Resource Operations	19
3.3.2 Training and Education	20
3.3.3 Core Research	20
4. Significance	22
5. Progress	30

5.1	Brief Statement of Prior Goals	30
5.1.1	Resource Operations	30
5.1.2	Training and Education	30
5.1.3	Core Research	31
5.2	Summary of Progress: 11/77 - 4/80	32
5.3	Detailed Progress Highlights	34
5.3.1	Resource Operations	34
5.3.1.1	System Hardware	34
5.3.1.2	System Software	39
5.3.1.3	Network Communication Facilities	46
5.3.1.4	Resource Management	47
5.3.2	Core Research	49
5.3.3	SUMEX Staff Publications	51
6.	Methods of Procedure	52
6.1	Resource Operations Plans	52
6.1.1	Resource Hardware	52
6.1.1.1	Rationale for Future Plans	52
6.1.1.2	Summary of Proposed Hardware Acquisitions	55
6.1.1.3	Existing Hardware Operation	56
6.1.1.4	Large Address Space Machines	57
6.1.1.5	Single-User Professional Workstations	57
6.1.1.6	File Server	59
6.1.2	Communication Networks	62
6.1.2.1	Long-Distance Connections	62
6.1.2.2	Local Intermachine Connections	62
6.1.3	Resource Software	64
6.1.4	Community Management	66

6.2	Training and Education Plans	68
6.3	Core Research Plans	69
6.3.1	Knowledge Representation	71
6.3.1.1	RLL -- The Representation Language Language	71
6.3.1.2	Research on Planning	73
6.3.1.3	Causal Models	77
6.3.2	Knowledge Utilization and Tools for Building Expert Systems	78
6.3.2.1	Attempt to Generalize (AGE)	78
6.3.2.2	AI Handbook	78
6.3.2.3	Research in Automated Consultation about Expert Systems	79
6.3.2.4	EMYCIN	80
6.3.3	Knowledge Acquisition	82
6.3.4	Explanation	85
7.	Available Facilities	89
8.	Literature Cited	90
9.	Collaborative Project Reports	135
9.1	Stanford Projects	136
9.1.1	AGE - Attempt to Generalize	137
9.1.2	AI Handbook Project	145
9.1.3	DENDRAL Project	149
9.1.4	MOLGEN Project	171
9.1.5	MYCIN Project	186
9.1.6	Protein Structure Project	205
9.1.7	RX Project	211

9.2	National AIM Projects	220
9.2.1	Acquisition of Cognitive Procedures (ACT)	221
9.2.2	SECS - Simulation and Evaluation of Chemical Synthesis	226
9.2.3	Hierarchical Models of Human Cognition	234
9.2.4	HMF - Higher Mental Functions	240
9.2.5	INTERNIST Project	244
9.2.6	PUFF/VM Project	250
9.2.7	Simulation of Cognitive Processes	261
9.2.8	Rutgers Computers in Biomedicine Project [Rutgers-AIM]	267
9.2.9	Decision Models in Clinical Diagnosis [Rutgers- AIM]	282
9.2.10	Heuristic Decisions in Metabolic Modeling [Rutgers-AIM]	286
9.3	Pilot Stanford Projects	289
9.3.1	Ultrasonic Imaging Project	290
9.4	Pilot AIM Projects	297
9.4.1	Coagulation Expert Project	298
9.4.2	Communication Enhancement Project	302
9.4.3	A Computerized Psychopharmacology Advisor	309
9.4.4	Computer-Aided Refinement of Medical Knowledge	317
9.4.5	Interactive Statistical Package Advisor	321
9.4.6	Conceptual Structures for Medical Diagnosis [Rutgers-AIM]	323

Appendix A	
Community Growth and Project Synopses	331
Appendix B	
Resource Operations and Usage Statistics	355
Appendix C	
Local Network Integration	374
Appendix D	
Remote Network Communication Facilities	376
Appendix E	
Resource Management Structure	383
Appendix F	
LISP Address Space Limitations	390
Appendix G	
AI Handbook Outline	392
Appendix H	
MAINSAIL System Demonstration	398
Appendix I	
AIM Management Committee Membership	399

List of Figures

Figure	Page
1. Current SUMEX-AIM KI-10 Computer Configuration	36
2. Current SUMEX-AIM 2020 Computer Configuration	37
3. Intermachine Connections via ETHERNET	38
4. Proposed VAX configuration	60
5. Planned Ethernet System to Integrate System Hardware	61
6. SUMEX-AIM Growth by Community	331
7. Total CPU Time Consumed by Month	356
8. Peak Number of Jobs by Month	357
9. Peak Load Average by Month	357
10. Monthly CPU Usage by Community	359
11. Monthly File Space Usage by Community	360
12. Monthly Terminal Connect Time by Community	361
13. Average Diurnal Loading (4/80): Number of Jobs	369
14. Average Diurnal Loading (4/80): Load Average	370
15. Average Diurnal Loading (4/80): Percent Time Used	370
16. TYMNET Terminal Connect Time	371

17.	ARPANET Terminal Connect Time	372
18.	TYMNET Network Node List	379
19.	ARPANET Geographical Network Map	380
20.	ARPANET Logical Network Map	381
21.	TELENET Geographical Network Map	382